



Research Report  
Statistical Research Unit  
Department of Economics  
Göteborg University  
Sweden

---

## Effect of dependency in systems for multivariate surveillance

**Eva Andersson**

**Research Report 2007:1**  
**ISSN 0349-8034**

---

Mailing address:	Fax	Phone	Home Page:
Statistical Research Unit P.O. Box 640 SE 405 30 Göteborg Sweden	Nat: 031-786 12 74	Nat: 031-786 00 00  Int: +46 31 786 12 74	<a href="http://www.statistics.gu.se/">http://www.statistics.gu.se/</a>

# Effect of dependency in systems for multivariate surveillance

E Andersson

*Statistical Research Unit, School of Economics and Commercial Law,  
Göteborg University, Sweden.*

---

## **Abstract**

In many situations we need a system for detecting changes early. Examples are early detection of disease outbreaks, of patients at risk and of financial instability. In influenza outbreaks, for example, we want to detect an increase in the number of cases. Important indicators might be the number of cases of influenza-like illness and pharmacy sales (e.g. aspirin). By continually monitoring these indicators, we can early detect a change in the process of interest. The methodology of statistical surveillance is used. Often, the conclusions about the process(es) of interest is improved if the surveillance is based on several indicators. Here three systems for multivariate surveillance are compared. One system, called LRpar, is based on parallel likelihood ratio methods, since the likelihood ratio has been shown to have several optimality properties. In LRpar, the marginal density of each indicator is monitored and an alarm is called as soon as one of the likelihood ratios exceeds its alarm limit. The LRpar is compared to an optimal alarm system, called LRjoint, which is derived from the full likelihood ratio for the joint density. The performances of LRpar and LRjoint are compared to a system where the Hotellings T2 is monitored. The evaluation is made using the delay of a motivated alarm, as a function of the times of the changes. The effect of dependency is investigated: both dependency between the monitored processes and correlation between the time points when the changes occur. When the first change occurs immediately, the three methods work rather similarly, for independent processes and zero correlation between the change times. But when all processes change later, the T2 has much longer delay than LRjoint and LRpar. This holds both when the processes are independent and when they have a positive covariance. When we assume a positive correlation between the change times, the LRjoint yields a shorter delay than LRpar when the changes actually do occur simultaneously, whereas the opposite is true when the changes do actually occur at different time point.

**Keywords:** Multivariate; Surveillance; Dependency; Optimal; Covariance; Likelihood ratio.

## 1. Introduction

In many situations it is important to monitor a process in order to detect an important change as soon as possible. Examples are turning point detection in business cycles (Neftci (1982), Hamilton (1989), Andersson et al. (2004)), in hormone cycles (Royston (1991)), in influenza cycles (Baron (2002)) and in financial cycles (Bock (2003)). Another area is detection of growth retardation of foetuses (Petzold et al. (2004)). Yet another is detection of an increased level, emerging from a source and spreading spatially (Järpe (1999)). In on-line monitoring we have repeated decisions: at each time point, a new observation becomes available and a new decision is made as to whether the process has changed or not. The methodology of statistical surveillance is appropriate.

Statistical surveillance is a methodology for discriminating between two events, "the change has occurred" and "the change has not occurred". The time of change is unknown. The time scale can differ between applications (days, weeks, months), but common to all surveillance are the repeated decisions, made at each time point. The decision is made using an alarm statistic and an alarm limit. In industrial quality control charts (e.g. xbar-charts), the decision rule can be that an alarm is given as soon as an observation crosses the alarm limit (Shewhart (1931)). There is always a risk for a false alarm, but the parameters of the surveillance method are chosen so that we know the false alarm property. For motivated alarms, i.e. when the change actually happens, we want a quick detection. Since we have repeated decisions, hence size and power are not appropriate measures. Instead we have a trade-off between false alarms and delay of motivated alarms or detection probability. The false alarms are often controlled by a fixed average run length,  $ARL^0$ . The delay can be measured by the expected delay time between the change time and the alarm. Besides the Shewhart method, the EWMA method (Roberts (1959)) and the CUSUM method (Page (1954)) can be mentioned.

In many situations we monitor several processes, which may change simultaneously or at different times. There are different approaches in multivariate monitoring: the data can be reduced to a scalar at each time or we can use separate alarm systems for each process in combination with an inference rule like union intersection. Also, multivariate versions of univariate methods have been suggested, for example MEWMA and MCUSUM. The choice of method depends on e.g. whether the processes under surveillance are correlated and whether the change times are dependent.

The aim of this paper is to compare different multivariate surveillance systems. In Section 1, optimal univariate surveillance is discussed and different approaches to multivariate surveillance are exemplified. This section also includes some previously suggested methods for multivariate surveillance. Section 2 presents the surveillance methods, used in this paper. In Section 3 and 4, the results of the study are presented and section 5 contains a discussion.

### *1.1 Optimal statistical surveillance*

Statistical surveillance is a methodology for discriminating between the two events C and D, such that C="the change has occurred" and D="the change has not occurred". By an

alarm system we decide whether D or C is the most likely. Optimal alarm systems are based on the likelihood ratio between C and D (Shiryayev (1963), Frisén and de Maré (1991)), see further in Section 1.2.3.

Statistical surveillance is used for on-line detection of an important change in the underlying process, for instance the expected value. For daily data, a new decision is made each day, based on the available data. When there is enough evidence of a change, an alarm is called. The surveillance system (alarm statistic and alarm limit) is specified to have a known false alarm risk by adjusting the alarm limit.

Generally, we get a new observation on the observed process X at each time  $s=\{1, 2, \dots\}$  and the surveillance system is used to discriminate between events C(s) and D(s). At an unknown time,  $\tau$ , there is a change in the distribution of X, so that

$$X(s) \sim \begin{cases} f(x(s)|D), & s < \tau \\ f(x(s)|C), & s \geq \tau \end{cases}$$

The events D and C are specified according to the change of interest.

As mentioned above, the likelihood ratio between C and D is optimal. If we want to detect a change at the current time s, then  $C=\{\tau=s\}$ ,  $D=\{\tau>s\}$  and the likelihood ratio is

$$\frac{f(x_s | \tau = s)}{f(x_s | \tau > s)} > k,$$

where  $x_s=\{x(1), x(2), \dots, x(s)\}$  and  $k$ =alarm limit. E.g. if it is of interest to detect a change in the expected value (from  $\mu^0$  to  $\mu^1$ ) and the observations are independent and normally distributed, then the optimal LR reduces to the last observation,  $x(s)$

$$\frac{f(x(s) | \mu = \mu^1)}{f(x(s) | \mu = \mu^0)} > k \iff x(s) > \frac{2\sigma^2 \cdot \ln(k) - (\mu^0)^2 - (\mu^1)^2}{2(\mu^1 - \mu^0)}.$$

This is the Shewhart method, used in e.g. an xbar-chart. Contrary to using only the latest observation, the CUSUM method cumulates the observations since the start of the surveillance. CUSUM is the maximal of the likelihood ratios at the decision time s. Another surveillance method is EWMA, where all observations are weighted together. EWMA can not be exactly derived from the likelihood ratio, but for certain values of the smoothing constant, EWMA is approximately the same as LR (see Frisén and Sonesson (2002)).

For the situation when it is important to detect if there has been a change at some time point since the start of the surveillance, we specify  $C = \{\tau \leq s\} = \{\{\tau=1\}, \{\tau=2\}, \dots, \{\tau=s\}\}$ . Then the LR consists of s components

$$w_1 \frac{f(x_s | \tau = 1)}{f(x_s | \tau > s)} + w_2 \frac{f(x_s | \tau = 2)}{f(x_s | \tau > s)} + \dots + w_s \frac{f(x_s | \tau = s)}{f(x_s | \tau > s)} > k_s, \quad (1)$$

where  $w_i=P(\tau=i)/P(\tau \leq s)$  and  $k_s = k \cdot P(\tau \geq s)/P(\tau < s)$ . Thus for  $C=\{\tau \leq s\}$ , it is optimal to use all observations  $\{x(1), \dots, x(s)\}$ . The time of alarm,  $t_A$ , is defined as the first time that the alarm statistic exceeds the alarm limit,

$$t_A = \min\{s: LR(s) > k_s\}.$$

When  $P(C) = 1 - P(D)$ , the likelihood ratio is equivalent to the posterior probability, with alarm rule

$$P(C|x_s) > k ,$$

where  $k$  is the constant alarm limit. The posterior probability is often used in hidden Markov model approaches (HMM).

The alarm limit is adjusted so that the false alarm property has a specified value. In quality control, a common false alarm property is the average run length to the first false alarm,  $ARL^0 = E[t_A | \tau = \infty]$ . Gan (1993) instead suggested the use of the median run length ( $MRL^0$ ). In many theoretical work, the false alarm probability is used, which summarize the false alarm distribution using the distribution of  $\tau$ , such that

$$PFA = \sum_{i=1}^{\infty} P(t_A < i) \cdot P(\tau = i).$$

For an on-line system, the ability to detect a change quickly is important, i.e. to have a short delay for motivated alarms. For most surveillance methods, the delay of an alarm depends on when the change do occur, in relation to the start of the surveillance and the delay is often longest when the change occurs at the start ( $\tau=1$ ). The conditional expected delay of an alarm (see Frisé and Wessman (1999)) is defined as

$$CED = E[t_A - \tau | t_A \geq \tau, \tau = t].$$

Many evaluations are made using  $\tau=1$ , e.g  $CED(1)$  which is equivalent to  $ARL^1 - 1$ . However it is important to consider other change point times also.

## 1.2 Different approaches to multivariate surveillance

In a multivariate setting there are  $p$  processes  $\{X_1, X_2, \dots, X_p\}$  for which

$$X_j(s) \sim \begin{cases} f(x_j(s)|D), & s < \tau_j \\ f(x_j(s)|C), & s \geq \tau_j \end{cases}$$

where  $j = \{1, 2, \dots, p\}$ .

An overview of multivariate surveillance is given in Sonesson and Frisé (2005), who categorize different approaches to multivariate surveillance into the following groups; reduction of dimensionality, reduction to one scalar statistic, parallel surveillance, vector accumulation and simultaneous solution. In this paper, three of these approaches are compared; reduction to scalar, parallel surveillance and simultaneous solution. A reduction to one scalar statistic at each time means that a summary statistic is calculated from the  $p$   $X$ -processes and that this statistic can be monitored by a system for univariate surveillance. In parallel surveillance, the marginal distribution of each variable is monitored with  $p$  parallel systems, from which the information is combined. In a simultaneous solution, the full  $p$ -variate joint distribution of the processes is used.

### 1.2.1 Reduction to one scalar statistic at each time

The  $p$  processes can be reduced to a scalar at each time, for example by calculating the (weighted) mean. Wessman (1998) showed that when the processes have identical change times ( $\tau_1=\tau_2=\dots=\tau_p=\tau$ ), then there exists a sufficient univariate reduction of the variables  $\{X_1, \dots, X_p\}$ . Thus, without loss of information, the multivariate data can be reduced to a scalar statistic and then univariate surveillance can be applied. The sufficiency also holds when the changes times are not identical but the lag times between them are known.

### 1.2.2 Parallel surveillance

Multivariate surveillance can be made by monitoring the marginal density of each process. One drawback is that no information about the dependency structure is used. The surveillance system for  $X_1$  is only concerned with  $\tau_1$ , and correspondingly for  $X_2, X_3$  etc. For process  $X_j$  at time  $s$ , we have alarm statistic  $p(x_{js})$  and alarm limit  $k_{js}$ , so that an alarm is called when

$$p(x_{js}) > k_{js},$$

where  $x_{js} = \{x_j(1), x_j(2), \dots, x_j(s)\}$ . The time of alarm of  $X_j$  is

$$t_{Aj} = \min t: p(x_{jt}) > k_{jt}.$$

The time of alarm for the whole system is defined as

$$t_A = \min \{t_{A1}, t_{A2}, \dots\}. \quad (2)$$

Does et al. (1999) uses separate surveillance for of each process (there each principal component) in a case study.

### 1.2.3 Simultaneous solution

It was shown by Shiryaev (1963) that the likelihood ratio between  $C$  and  $D$  is optimal in the sense that is maximized the expected utility,  $E[u]$ , where  $u$  equals

$$u(t_A, \tau) = \begin{cases} h(t_A - \tau) & , t_A < \tau \\ a_1 \cdot (t_A - \tau) + a_2 & , t_A \geq \tau \end{cases}$$

The optimality was proven for the situation when  $\tau$  follows a Geometric distribution (with parameter  $\nu$ ). The utility expression above is easily applied to the univariate situation. If the function  $h(t_A - \tau)$  is a constant,  $b$ , then it has been shown that the utility is maximized when the delay,  $t_A - \tau$ , is minimized (see Frisén (2003)). The minimal expected delay was shown to hold also for a situation where  $\tau$  is not Geometrically distributed (Andersson (2004)).

For multivariate data,  $\{X_1, X_2, \dots, X_p\}$ , there can be  $p$  change points ( $\tau_1, \tau_2, \dots, \tau_p$ ). We study the situation when we want to detect the first change, i.e. we want to make inference about  $\tau_{(1)} = \min[\tau_1, \tau_2, \dots, \tau_p]$ .

Then the utility can be written as

$$u(t_A, \tau_{(1)}) = \begin{cases} h(t_A - \tau_{(1)}) & , t_A < \tau_{(1)} \\ a_1 \cdot (t_A - \tau_{(1)}) + a_2, & t_A \geq \tau_{(1)} \end{cases}$$

where  $t_A$  is the time of alarm for the whole system (e.g. (2) in Section 1.2.2). The distribution of  $\tau_{(1)}$  depends on the p-variate distribution for  $(\tau_1, \dots, \tau_p)$ . For  $p=2$ , we use the bivariate Geometric distribution, see e.g. Marshall and Olkin (1997). Then  $\tau_{(1)}$  follows a Geometric distribution (Sun and Basu (1995)) and the optimality holds: the likelihood ratio maximizes the expected utility.

Wessman (1999) investigated the situation where it is of interest to decide whether the first change has occurred at the current time point,  $C=\{\tau_{(1)}=s\}$ . In this paper we want to decide whether the first change has started at any time point since the beginning of the surveillance,  $C=\{\tau_{(1)} \leq s\} = \{\{\tau_{(1)}=1\}, \dots, \{\tau_{(1)}=s\}\}$ . The D event is specified as  $\{\tau_{(1)} > s\}$ . The optimal surveillance system (simultaneous solution) is derived from the likelihood ratio. Since  $P(C)=1-P(D)$ , the likelihood ratio is equivalent to the posterior probability

$$P(C|m_s) > k,$$

where  $m_s = \{m(1), \dots, m(s)\} = \{\{x_1(1), \dots, x_p(1)\}, \dots, \{x_1(s), \dots, x_p(s)\}\}$  and  $k = \text{constant alarm limit}$ . Since  $P(C) = 1 - P(D)$ , the alarm rule is

$$\frac{P(m_s \cap C)}{P(m_s \cap C) + P(m_s \cap D)} > k.$$

Hence forward we express the likelihood function by  $f(*)$  instead of  $P(*)$ , so that

$$\frac{f(m_s \cap C)}{f(m_s \cap D)} > \frac{k}{1-k}$$

We specify  $C = \{\tau_{(1)} \leq s\}$  and  $D = \{\tau_{(1)} > s\}$  and the likelihood ratio alarm rule consists of partial likelihood ratios

$$w_s^1 \cdot \frac{f(m_s | \tau_{(1)} = 1)}{f(m_s | \tau_{(1)} > s)} + w_s^2 \cdot \frac{f(m_s | \tau_{(1)} = 2)}{f(m_s | \tau_{(1)} > s)} + \dots + w_s^s \cdot \frac{f(m_s | \tau_{(1)} = s)}{f(m_s | \tau_{(1)} > s)} > k_s,$$

expressed as

$$\sum_{j=1}^s w_s^j \cdot L(s, j).$$

The weights,  $w_s^t = P(\tau_{(1)}=t)/P(\tau_{(1)} \leq s)$ , and the limit,  $k_s = k \cdot P(\tau_{(1)} > s)/P(\tau_{(1)} \leq s)$ , depends on the distribution of  $\tau_{(1)}$ , which in turn depends on the joint distribution of  $(\tau_1, \dots, \tau_p)$ . The time of alarm is defined as

$$t_A = \min t: LR(m_s) > k_s.$$

### *1.3 Suggested approaches for multivariate surveillance*

Some of the suggested approaches for monitoring multivariate data are reviewed.

#### 1.3.1 Multivariate monitoring using a reduction

As mentioned before, Wessman (1998) showed that when the changes occur simultaneously, there exists a sufficient reduction of the data. Multivariate data can e.g. be reduced by the Hotellings T2 statistic or by principal component analysis.

When the changes occur at different time points, a problem with a reduction is to determine which variable that causes the alarm. Javaheri and Houshmand (2001) suggest a follow-up with discriminant analysis to investigate which variables that cause the alarm. Jolayemi (2000) constructs multiple control regions for two assignable causes. Kalagonda and Kulkarni (2003) propose constructing a model for the in-control process and using dummy variables to determine the nature of the change. Mason et al. (1995) decompose T2 into independent components, each reflecting an individual variable  $X_j$ .

Kourti and MacGregor (1996) show that the T2 reduction works well when the dimension is not too high.

Wikström et al. (1998) use principal component analysis to derive the most important principal components and then univariate CUSUM and univariate EWMA is applied to the first principal component.

Abu-Shawiesh and Abdullah (2001) study surveillance of two correlated processes. The T2 statistic is based on robust estimates of location and scale and the situation is one where both scale and location change. This approach is compared to the ordinary T2.

Mason et al. (2003) find that AR processes result in a U shaped T2 curve and if the data contains trend, the T2 values will exhibit a trend.

Another reduction is to use the minimum and maximum values at each time, as is done in Sepúlveda and Nachlas (1997). Charnes (1995) studies the situation where the processes under surveillance are time dependent and where the residuals (or forecasting errors) are used in the monitoring. Kang and Albin (2000) model the variable Y as a function of the variable X and monitor the slope and intercept by T2 statistic as well as the residuals (deviations from reference line).

Aparisi et al. (2001) reduce the covariance matrix through the determinant or the trace. This scalar is monitored, using univariate Shewhart and univariate EWMA and CUSUM. Guerrero-Cusumano (1995) uses an entropy measure instead of the determinant.

Stoumbos and Jones (2000) reduce the multivariate data to a probability measure which shows how “central” the observation is.

Lu et al. (1998) study correlated variables and the proportion of non-conforming units. The multivariate data are reduced to a scalar by weighting together the non-conforming units and then monitored by univariate Shewhart.

Cheng and Liu (2000) use a rank measure for how outlying an observation is and then the univariate rank variable is monitored using a Shewhart approach.

Koskinen and Öller (2004) suggest the use of a weighted index, which is monitored using univariate methods. A similar approach is used in Talluri and Sarkis (2002).



Qiu and Hawkins (2001) use the anti-ranks of the  $p$  observations at each time and summarizes the anti ranks in a statistic which is monitored using univariate CUSUM.

### 1.3.2 MEWMA and MCUSUM

From the univariate specification of the EWMA and CUSUM methods, several multivariate variants have been developed.

Bodden and Rigdon (1999) use a multivariate EWMA and smooth all  $p$  processes using the same constant,  $\lambda$ , and then the vector of the smoothed values is reduced by the T2 statistic. This approach is also used in Love and Linderman (2003) and Molnau et al. (2001) and by Stoumbos and Sullivan (2002), who show that if the smoothing constant equals 1 (no smoothing) and the process is not normally distributed, then the  $ARL^0$  is over estimated if the alarm limits are determined under normality assumptions. But for a small smoothing constant, the  $ARL^0$  is not so biased, if the normality assumption is violated. A small smoothing constant gives approximately equal weight to all observations, which is approximately normal according to the central limit theorem.

Lowry et al. (1992) use EWMA smoothing with separate  $\lambda$  values, and then reduction by the T2 statistic. The  $\lambda$  values are determined so as to minimize the  $ARL^1$ . Yumin (1996) suggests that if the  $p$   $X$  processes are correlated, they should be transformed into principal components, which are then smoothed separately. Runger et al. (1999) reduce the dimension by a transform similar to principal component analysis, then the transformations are smoothed using the same  $\lambda$  and then the T2 statistic is calculated from the smoothed series. Gan (1997) constructs a control chart with the smoothed variance on one axis and the smoothed mean on the other, called a combined EWMA chart. The advantage, to a T2 reduction, is that the chart shows whether it is the mean or the variance that is out-of control.

A similar approach is used in Woodall and Ncube (1985), but with a univariate CUSUM for each of the  $p$  processes. Hawkins (1991) applied a univariate CUSUM to a linear combination of the variables  $\{X_1, X_2, \dots, X_p\}$ . Wessman (1998) showed that if the  $p$  processes have identical change times, then data can, without loss of information, be reduced to a univariate index.

Crosier (1988) and Pignatiello and Runger (1990) used an MCUSUM on the same form as as the univariate CUSUM, only with matrixes instead of scalars.

## 2. Model and methods

In this paper we study the situation when two processes may change at different time points, i.e.  $\tau_1$  and  $\tau_2$  may have different values. The  $\tau$  values can be dependent.

The aim is to detect an increase in the expected value, for example increases in the expected number of confirmed influenza cases and expected number of patients with influenza-like illness (Swedish institute of infectious disease, SMI).

At decision time  $s$ , the observations  $(X, Y)$  are modeled according to

$$\begin{pmatrix} m(1) \\ m(2) \\ \dots \\ m(s) \end{pmatrix} = \begin{pmatrix} x(1) & y(1) \\ x(2) & y(2) \\ \dots & \dots \\ x(s) & y(s) \end{pmatrix} = \begin{pmatrix} \mu_x(1) & \mu_y(1) \\ \mu_x(2) & \mu_y(2) \\ \dots & \dots \\ \mu_x(s) & \mu_y(s) \end{pmatrix} + \begin{pmatrix} \varepsilon_x(1) & \varepsilon_y(1) \\ \varepsilon_x(2) & \varepsilon_y(2) \\ \dots & \dots \\ \varepsilon_x(s) & \varepsilon_y(s) \end{pmatrix} \quad (4)$$

The stochastic term at time  $t$ ,  $(\varepsilon_x(t) \ \varepsilon_y(t))$ , follows a bivariate normal distribution, with expected value zero and covariance matrix

$$\sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad (5)$$

where  $\rho$ =correlation between  $X(t)$  and  $Y(t)$ .

The variables  $X(t)$  and  $Y(t)$  are (possibly) dependent but not  $X(t)$  and  $X(t-j)$ ,  $Y(t)$  and  $Y(t-j)$  or  $X(t)$  and  $Y(t-j)$ . The aim is to detect the first change in either of the  $\mu$  vectors  $(\mu_x, \mu_y)$ , where one application is detection of the start of an influenza epidemic. The SMI receives weekly information about confirmed cases (laboratory diagnosed, see Figure 1) and suspected cases. The development is roughly captured through the following parametric model for  $\mu$ , which will be used for both  $X$  and  $Y$  in the investigation below.

$$\mu(t): \begin{cases} \mu^D(t): \beta_{01}, & t < \tau \\ \mu^{Cr}(t): \beta_{02} + \beta_1 \cdot (t+1-\tau) + \beta_2 \cdot (t+1-\tau)^3, & t \geq \tau \end{cases} \quad (6)$$

where  $\beta_1 < 0$ ,  $\beta_2 > 0$ . Thus, for  $\tau_X = \tau_Y$ ,  $X$  and  $Y$  have the same distribution and they are independent (or dependent) conditional on  $\tau$ . The model above might be too crude in a real situation, but is used here to demonstrate the inferential aspects.

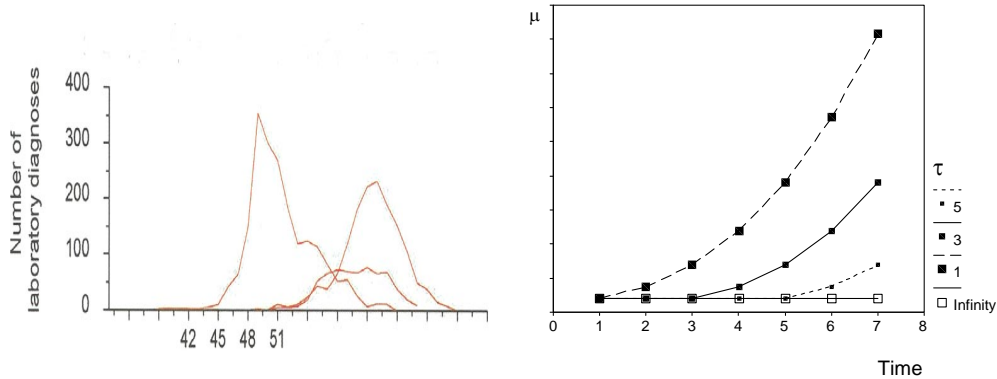


Figure 1: Left: The number of confirmed cases of influenza in Sweden for three consecutive seasons (source: SMI). Right: The model for  $\mu$  (for different starting times).

As mentioned above, a bivariate Geometric distribution is used for  $(\tau_X, \tau_Y)$ , with parameters  $(v_{01}, v_{10}, v_{11})$ . The marginal distributions are  $\tau_X \sim \text{Geo}(v_{1.} = v_{10} + v_{11})$  and  $\tau_Y \sim \text{Geo}(v_{.1} = v_{01} + v_{11})$ . The two marginal intensities are assumed to be equal

( $v_{11}=v_{.11}=0.10$ ). The value of  $v_{11}$  will be varied, thereby yielding different correlations, here  $\{0, 0.80\}$ . The correlation between  $\tau_X$  and  $\tau_Y$  is from now denoted  $\psi$ .

Three methods will be compared, corresponding to the approaches of reduction to a scalar, parallel surveillance and simultaneous solution.

### 2.1 The T2 method in surveillance

An early multivariate method is the T2 method of Hotelling (1947). The statistic,  $(\underline{x}-\underline{\mu}^D)' \Sigma^{-1} (\underline{x}-\underline{\mu}^D)$ , is here assumed to have a known covariance matrix, see e.g. Alt (1985). The T2 is an example of a reduction. We assume that X and Y have the same distribution, conditional on  $\tau_X$  and  $\tau_Y$  and thus an alarm is given when

$$T2(s) = \frac{(x(s)-\mu^D(s))^2}{(1-\rho^2)\sigma^2} + \frac{(y(s)-\mu^D(s))^2}{(1-\rho^2)\sigma^2} - \frac{2\rho(x(s)-\mu^D(s))(y(s)-\mu^D(s))}{(1-\rho^2)\sigma^2} > k.$$

### 2.2 Parallel likelihood ratio systems

The parallel surveillance system in this paper is based on separate likelihood ratio methods for X and Y. This system is hence forward denoted LRpar. For the process X we have the following likelihood ratio alarm statistic

$$LR_X(s) = \sum_{j=1}^s \frac{P(\tau = j)}{P(\tau \leq s)} \exp \left( \frac{\sum_{t=1}^s (x(t) - \mu^D(t))^2}{2\sigma^2} - \frac{\sum_{t=1}^s (x(t) - \mu^{Cj}(t))^2}{2\sigma^2} \right) > k \frac{P(\tau_X > s)}{P(\tau_X \leq s)}$$

where  $\mu^D$  and  $\mu^C$  are defined in (6) and  $\tau$  is assumed to follow a geometric distribution. The  $LR_Y(s)$  statistic is defined correspondingly. The time of alarm for X is defined as  $t_{AX} = \min[LR_X(s) > k_s^X]$

where  $k_s^X = k \cdot (P(\tau_X > s)/P(\tau_X \leq s))$ , and correspondingly for Y. The time of alarm for the LRpar system is the first time for which either alarm system gives an alarm, i.e.

$$t_A = \min\{t_{AX}, t_{AY}\}.$$

If X and Y are independent, the distribution of  $t_A$  is a direct function of the distributions of  $t_{AX}$  and  $t_{AY}$ .

The two alarm limits,  $k_s^X$  and  $k_s^Y$ , are adjusted to yield a specified false alarm property for the whole system. If there is no particular information regarding the cost for false alarms, then it is natural to have the same false alarm property for the two methods (e.g. equal  $ARL^0$ ). This approach is used in this paper since we assume that X and Y have the same distribution, conditional on  $\tau_X$  and  $\tau_Y$ . If inspection and restoration costs are not the same for the processes, Serel et al. (2000) suggest to have different type I errors.

### 2.3 Simultaneous solution using the joint likelihood ratio

The optimal surveillance method for  $C = \{\tau_{(1)} \leq s\}$  was derived in Section 1.2.3 and was shown to consist of  $s$  weighted partial likelihood ratios  $\{L(s,1), \dots, L(s,s)\}$ , such that

$$w_s^j \cdot L(s, j) = \frac{P(\tau_{(1)} = j)}{P(\tau_{(1)} \leq s)} \cdot \frac{f(m_s | \tau_{(1)} = j)}{f(m_s | \tau_{(1)} > s)}.$$

The event  $\{\tau_{(1)} = j\}$  consists of the following sub-events:

- $\{\tau_X = j \cap \tau_Y > j\}$ , event C-D, “a turn in X at time j”
- $\{\tau_X > j \cap \tau_Y = j\}$ , event D-C, “a turn in Y at time j”
- $\{\tau_X = j \cap \tau_Y = j\}$ , event C-C, “turn in both X and Y at time j”.

Thus,  $w_s^j \cdot L(s, j)$  is weighted together as

$$w_s^{j0} \cdot L(s, j, C-D) + w_s^{0j} \cdot L(s, j, D-C) + w_s^{jj} \cdot L(s, j, C-C) =$$

$$w_s^{j0} \cdot \frac{f(m_s | \tau_X = j, \tau_Y > j)}{f(m_s | \tau_X > j, \tau_Y > j)} + w_s^{0j} \cdot \frac{f(m_s | \tau_X > j, \tau_Y = j)}{f(m_s | \tau_X > j, \tau_Y > j)} + w_s^{jj} \cdot \frac{f(m_s | \tau_X = j, \tau_Y = j)}{f(m_s | \tau_X > j, \tau_Y > j)}$$

where the weights are

$$w_s^{j0} = \frac{P(\tau_X = j \cap \tau_Y > j)}{P(\tau_{(1)} \leq s)}, \quad w_s^{0j} = \frac{P(\tau_X > j \cap \tau_Y = j)}{P(\tau_{(1)} \leq s)}, \quad w_s^{jj} = \frac{P(\tau_1 = j \cap \tau_2 = j)}{P(\tau_{(1)} \leq s)}.$$

Thus, the information about the bivariate distribution of  $(\tau_X, \tau_Y)$  is included through the weights and the alarm limit,  $k' P(\tau_{(1)} > s) / P(\tau_{(1)} \leq s)$ .

The complete optimal LR statistic at decision time  $s$  consists of  $3 \cdot s$  components

$$\sum_{j=1}^s \left[ w_s^{j0} \frac{f(m_s | \tau_X = j, \tau_Y > j)}{f(m_s | \tau_X > j, \tau_Y > j)} + w_s^{0j} \frac{f(m_s | \tau_X > j, \tau_Y = j)}{f(m_s | \tau_X > j, \tau_Y > j)} + w_s^{jj} \frac{f(m_s | \tau_X = j, \tau_Y = j)}{f(m_s | \tau_X > j, \tau_Y > j)} \right].$$

For example, the event  $\{\tau_X = j, \tau_Y > j\}$  means that there is a change in  $\mu^X$  at time  $j$ , but no change yet in  $\mu^Y$ . Then the  $L(s, j)$  equals

$$w_s^{j0} \frac{f(x_s, y_s | \mu_X = \mu^{Cj}, \mu_Y = \mu^D)}{f(x_s, y_s | \mu_X = \mu^D, \mu_Y = \mu^D)} + w_s^{0j} \frac{f(x_s, y_s | \mu_X = \mu^D, \mu_Y = \mu^{Cj})}{f(x_s, y_s | \mu_X = \mu^D, \mu_Y = \mu^D)} +$$

$$w_s^{jj} \frac{f(x_s, y_s | \mu_X = \mu^{Cj}, \mu_Y = \mu^{Cj})}{f(x_s, y_s | \mu_X = \mu^D, \mu_Y = \mu^D)},$$

where  $f(x_s, y_s)$  is the bivariate normal distribution with expected values as in (4) and covariance matrix as in (5). Because of independence over time, we have e.g.

$$L(s, j, C-C) = \frac{f(x_s, y_s | \mu^X = \mu^{Cj}, \mu^Y = \mu^{Cj})}{f(x_s, y_s | \mu^X = \mu^D, \mu^Y = \mu^D)} = \prod_{j=1}^s \frac{f(x(j), y(j) | \mu^X = \mu^{Cj}, \mu^Y = \mu^{Cj})}{f(x(j), y(j) | \mu^X = \mu^D, \mu^Y = \mu^D)}.$$

For example, the denominator,  $f(x_s, y_s | \mu^X = \mu^D, \mu^Y = \mu^D)$ , is

$$c_1 \cdot \exp \left( -c_2 \cdot \sum_{t=1}^s \left[ \left( \frac{x(t) - \mu^D(t)}{\sigma_X} \right)^2 - 2\rho \left( \frac{x(t) - \mu^D(t)}{\sigma_X} \right) \left( \frac{y(t) - \mu^D(t)}{\sigma_Y} \right) + \left( \frac{y(t) - \mu^D(t)}{\sigma_Y} \right)^2 \right] \right)$$

where  $c_2 = \frac{1}{2 \cdot (1 - \rho^2)}$ ,  $c_1 = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$ .

This system is an example of a simultaneous solution where the joint distribution is used, both for  $(\tau_X, \tau_Y)$  and for  $(X, Y)$ . The system is hence forward denoted the LRjoint. If X and Y are independent, the L(s, j) simplifies to

$$w_s^{j0} \frac{f(x_s | \tau_X = j)}{f(x_s | \tau_X > s)} + w_s^{0j} \frac{f(y_s | \tau_Y = j)}{f(y_s | \tau_Y > s)} + w_s^{jj} \frac{f(x_s | \tau_X = j)}{f(x_s | \tau_X > s)} \cdot \frac{f(y_s | \tau_Y = j)}{f(y_s | \tau_Y > s)}$$

Thus, for independent X and Y, the components L(s,j,C-D) and L(s,j, D-C) are the same components that are used in the LRpar.

## 2.4 Alarm regions for the methods

The methods T2, LRpar and LRjoint are made comparable by adjusting their respective alarm limits so that they all have the false alarm probability 0.10, defined as

$$\text{PFA} = \sum_{i=1}^{\infty} P(t_A < i) \cdot P(\tau_{(1)} = i). \quad (3)$$

The alarm regions for the three methods are illustrated in Figure 2, for different values of  $\psi$  (correlation between the change times) and different values of  $\rho$  (dependency between X and Y).

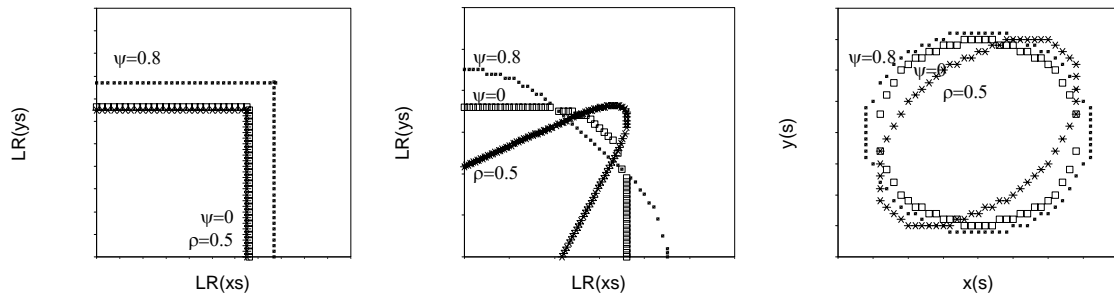


Figure 2: Alarm limit is marked by open square for  $\{\psi=0, \rho=0\}$ , filled square for  $\{\psi=0.8, \rho=0\}$ , star for  $\{\rho=0.5, \psi=0\}$ . Left: LRpar, middle: LRjoint, right: T2.

For LRpar, an alarm is called if at least one of the two LR statistics exceeds its alarm limit, thus if either X or Y deviates largely from the D-state. The LRjoint consists of three components, corresponding to changes in either  $\mu^X$  or  $\mu^Y$  or both. When X and Y are independent, LRjoint calls an alarm if at least one of the LR statistic are high or if the

joint LR statistic is high, thus if either X or Y or (X+Y) deviates from the D-state. The same holds for T2.

For LRpar, the shape of the alarm region (square) is unaffected by both  $\psi$  and  $\rho$ . For LRjoint, the shape changes in one direction when  $\psi$  increases and in another direction when  $\rho$  increases. The T2 method has a circle shaped alarm region for independent X and Y, whose size changes with  $\psi$ . But when  $\rho$  increases, the shape of the alarm region becomes more tilted and oval.

### 2.5 Evaluation in multivariate surveillance

The aim is to make inference about the first change,  $\tau_{(1)}$ , and the false alarms are controlled by PFA (see (3)). The performance in surveillance can be measured by the delay of a motivated alarm, e.g. by the CED in Section 1.1. In multivariate surveillance, the delay depends on both change times,  $\tau_X$  and  $\tau_Y$ , and the CED can be defined as

$$\text{CED}(t_1, t_2) = E[t_A - \tau_{(1)} \mid t_A \geq \tau_{(1)}, \tau_X = t_1, \tau_Y = t_2]. \quad (7)$$

### 3. Results on false alarms

The methods T2, LRpar and LRjoint are compared for different types of dependencies. The methods are comparable by all having PFA=0.1, where PFA is summarized using the bivariate Geometric distribution that results in a specified  $\psi$  (correlation between  $\tau_X$  and  $\tau_Y$ ). In LRpar, the intensity  $\nu=0.1$  is used. LRjoint includes the intensity parameters  $\nu_{11}=0.1$ ,  $\nu_{12}=0.1$  and  $\nu_{21}$ , where  $\nu_{11}$  takes different values for different  $\psi$ 's.

Table 1:  $\text{MRL}^0$  (median run length), for PFA=0.10

	$\rho(X, Y) = 0$		$\rho(X, Y) = 0.5$
	$\psi(\tau_X, \tau_Y) = 0$	$\psi(\tau_X, \tau_Y) = 0.8$	$\psi(\tau_X, \tau_Y) = 0$
LRjoint	14 ( $\nu_{11}=0.01$ )	36 ( $\nu_{11}=0.0895$ )	14 ( $\nu_{11}=0.01$ )
LRpar	14	34	14
T2	26	50	26

For a specific situation in Table 1, LRpar and LRjoint have similar  $\text{MRL}^0$ , whereas T2 has a higher median, because of the many early alarms (see Figure 3).

For all methods,  $\text{MRL}^0$  is almost independent on  $\rho$  (dependency between X and Y), but very dependent on  $\psi$  (correlation between  $\tau_X$  and  $\tau_Y$ ). The reason is that the density of  $\tau_{(1)}$  is used to weight together the alarm times. As  $\psi$  increases towards 1.00, the alarms tend to be more uniformly distributed and  $\text{MRL}^0$  is longer (see Figure 3).

The  $\text{MRL}^0$  increases when  $\psi=0.8$  is assumed. For LRpar, the alarm limit needs to be changed when a different  $\tau_{(1)}$ -density is used to summarize the false alarm distribution (Figure 4). This also holds for T2. For LRjoint, the alarm statistic itself is changed when  $\psi=0.8$ , since the three components C-D, D-C and C-C are weighted differently (see Section 2.3) and this affects the false alarm distribution.

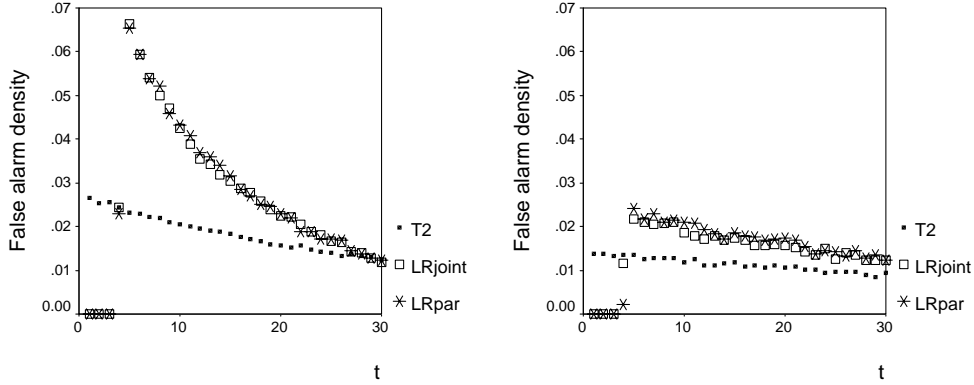


Figure 3: False alarm density,  $P(t_A=t)$ . Left:  $\psi=0$ , right:  $\psi=0.8$ .

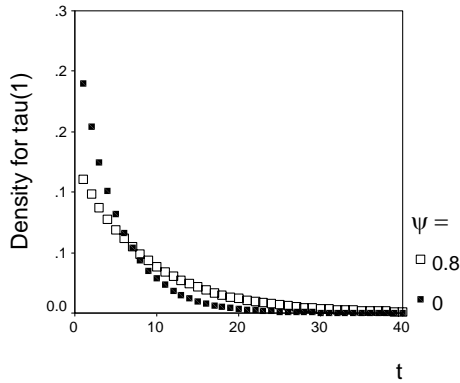


Figure 4: Probability density for  $\tau_{(1)}$ , when  $\psi=\{0,0.8\}$ .

#### 4. Using information about correlation between change times

Situations where  $X$  and  $Y$  may change at different time points,  $\tau_X$  and  $\tau_Y$ , are considered. The delay of a motivated alarm, in relation to the first change point, is used to compare the methods (CED( $t_1, t_2$ ) in (7)). The delay is studied for  $\psi=\{0, 0.8\}$ , i.e. when we assume that the changes occur independently and when they tend to occur more simultaneously.

##### 4.1 Comparison between the three methods

The effect of assumptions regarding the correlation between  $\tau_X$  and  $\tau_Y$  is investigated. The LRjoint is the only method (of these three) that uses the information of  $\psi$  in the alarm statistic, by different weights for the three components (see Section 2.3).

Results on CED( $t_1, t_2$ ) are presented for two situations, the first of which is when the first change occurs immediately,  $\tau_{(1)}=1$ . The second situation concerns later changes: for LRjoint and LRpar, the CED-pattern stabilizes after a while, e.g. CED(10,  $t$ ) has the same pattern as CED(14,  $t$ ),  $t=\{1, 2, \dots, 14\}$ , see Figure 5. For T2 the pattern is stabilized from the start. The stabilized pattern for CED( $t_1, t_2$ ) is here denoted steady curve.

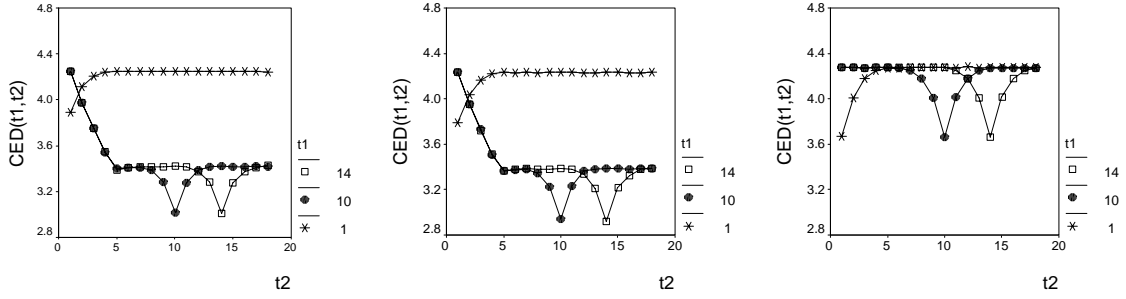


Figure 5:  $CED(t_1, t_2)$ , for  $t_1=\{1, 10, 14\}$ , where the pattern is the same for  $CED(10, t_2)$  and  $CED(14, t_2)$ . This is called steady curve. Left: LRpar, middle: LRjoint, right: T2.

#### 4.1.1 First change occurs immediately

The  $CED(t_1, t_2)$  is compared between the methods for  $\psi=\{0, 0.8\}$ , when the first change occurs immediately, i.e.  $\tau_{(1)}=1$ .

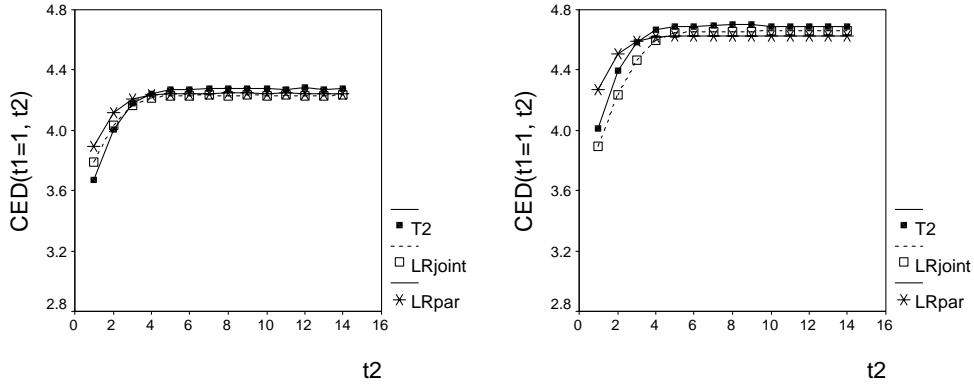


Figure 6:  $CED(t_1, t_2)$  when  $t_1=1$  and  $t_2=\{1, 2, \dots, 14\}$ . Left:  $\psi=0$ , right:  $\psi=0.8$ .

For all methods the delay is shortest when both changes occur immediately ( $\tau_X=\tau_Y=1$ ), independent of  $\psi$ .

When independent change times are assumed ( $\psi=0$ ), the methods give similar delay. For LRjoint and LRpar, this is also indicated by the similar alarm regions in Figure 2. The exception is at  $(\tau_X = \tau_Y = 1)$  where LRjoint has slightly shorter delay. LRjoint consists of three components (C-D, D-C, C-C, see section 2.3), whereas LRpar is based on two components (C-D and D-C). For  $\psi=0$ , T2 has the shortest delay.

When a positive correlation is assumed between the change times ( $\psi=0.8$ ), LRjoint is superior at  $(\tau_X = \tau_Y = 1)$ . LRjoint assigns a large weight to the C-C component and when the changes actually occur simultaneously, LRjoint has a short delay.

#### 4.1.2 Steady curve



In Figure 7, the  $CED(t_1, t_2)$  of the steady curve is compared between the methods for  $\psi=\{0, 0.8\}$ .

When one change occurs late (here  $t_1=10$ ), the methods are more different than for  $t_1=1$ . For  $\psi=0$ , T2 has the longest CED, except at  $(t_1=10, t_2=1)$ . The CED curves for LRpar and LRjoint are rather similar.

Also for  $\psi=0.8$ , T2 has longer CED than the two LR methods. For  $\psi>0$ , the difference is larger between LRpar and LRjoint: at simultaneous change times ( $\tau_X = \tau_Y$ ), LRjoint has shortest delay, but for  $\tau_X \neq \tau_Y$ , LRpar is a little better. LRjoint assigns a large weight to the C-C-component, which is an advantage when  $\tau_X = \tau_Y$ , but not when  $\tau_X \neq \tau_Y$ .

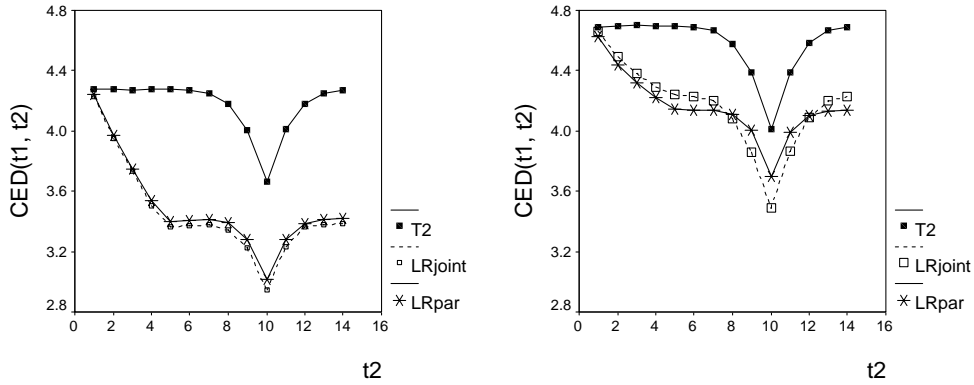


Figure 7:  $CED(t_1, t_2)$  when  $t_1=10$  and  $t_2 = \{1, 2, \dots, 14\}$ . Left:  $\psi=0$ , right:  $\psi=0.8$ .

#### 4.2 Effect of change times

As mentioned above, the delay depends on the change times,  $\tau_X$  and  $\tau_Y$ . The effect of assumptions about the correlation between  $\tau_X$  and  $\tau_Y$  is studied for each method.

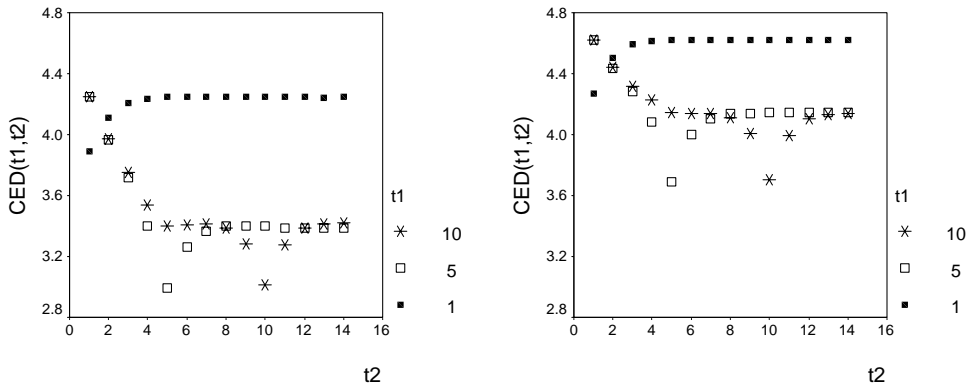


Figure 8a:  $CED(t_1, t_2)$  for LRpar when  $t_1 = \{1, 5, 10\}$ . Left:  $\psi=0$ , right:  $\psi=0.8$ .

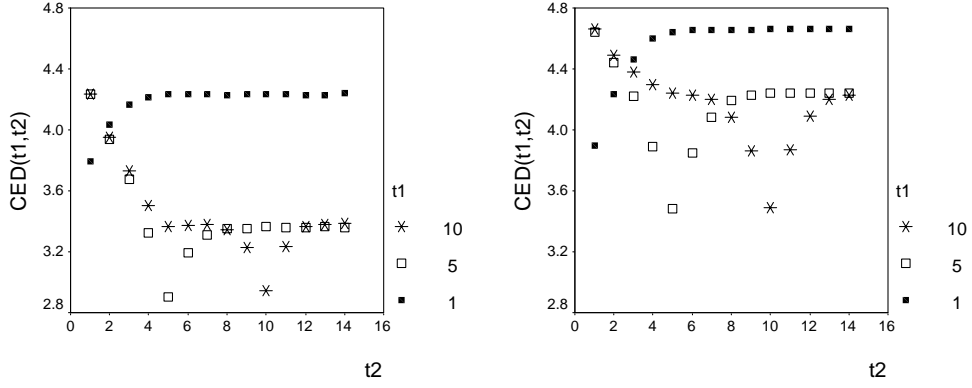


Figure 8b:  $CED(t_1, t_2)$  for LRjoint when  $t_1 = \{1, 5, 10\}$ . Left:  $\psi=0$ , right:  $\psi=0.8$ .

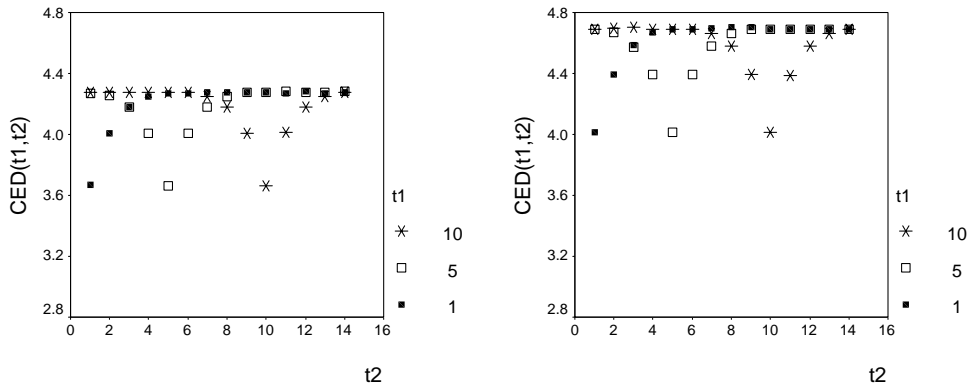


Figure 8c:  $CED(t_1, t_2)$  for T2 when  $t_1 = \{1, 5, 10\}$ . Left:  $\psi=0$ , right:  $\psi=0.8$ .

From Figure 8 we conclude that for each value of  $t_1$ , the  $CED(t_1, t_2)$  curve reaches an asymptote after a while,  $CED(t_1, \infty)$ . For T2, this asymptote is independent of  $t_1$ , whereas for LRpar and LRjoint it depends on  $t_1$ , for small values of  $t_1$ .

For all methods, CED is longer when  $\psi > 0$ . For LRpar, the CED-curve, for each value of  $t_1$ , is higher as a result of the higher alarm limit. The same holds for T2. Also for LRjoint, the CED values are higher, but there is also a larger difference between  $CED(t, t)$  and  $CED(t, \infty)$ . As discussed above, for  $\psi > 0$  the LRjoint works well when the changes do actually occur simultaneously.

Both for  $\psi=0$  and  $\psi=0.8$ , simultaneous changes have shortest delay for each method, i.e.  $CED(t, t) < CED(t, j)$ ,  $j \neq t$ . For  $CED(t, t)$ , both X and Y are in C-state at each time point.

For the two LR methods, simultaneous later changes have shortest delay, because then more data are available at the change times (e.g.  $CED(1, 1) > CED(5, 5)$ ). But T2 has the same delay for e.g. ( $\tau_X = \tau_Y = 1$ ) as for ( $\tau_X = \tau_Y = 5$ ), since T2 does only use observations from the current time point, thus T2 is not based on more data at later time points.

## 5. Processes under surveillance are dependent

The methods are compared when the processes X and Y are independent or have a positive covariance,  $\rho=\{0, 0.5\}$ . No correlation is assumed between the change times ( $\psi=0$ ).

### 5.1 Comparison between the three methods

For T2 and LRjoint, the alarm statistic does incorporate the information of  $\rho$ , whereas the alarm statistic of LRpar is the same, independent of  $\rho$ .

#### 5.1.1 First change occurs immediately

The CED curves, when  $\tau_{(1)}=1$ , are compared between the methods, for  $\rho=\{0, 0.5\}$ .

For immediate changes in both processes ( $\tau_X=\tau_Y=1$ ), the T2 has the shortest CED and LRpar the longest, when X and Y are independent ( $\rho=0$ ). However, for  $\rho=0.5$ , LRpar has shortest delay and LRjoint longest. One reason for the long delay of LRjoint is the assumption that  $\psi=0$ . Also, Wessman (1999) investigated multivariate surveillance for the situation  $C=\{\tau_{(1)}=s\}$  and showed that if changes occur in all processes, the probability of detecting it is lower if the observations are highly correlated. This agrees with the long CED for T2 and LRjoint for  $\tau_X=\tau_Y=1$  (see section 5.2.1).

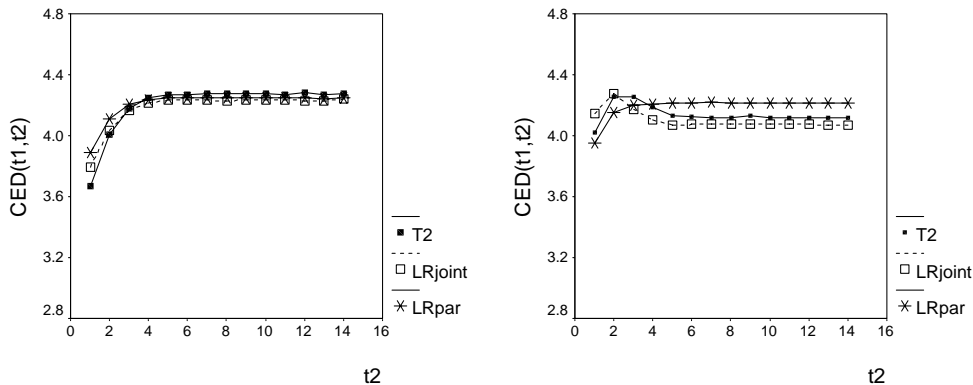


Figure 9:  $CED(t_1, t_2)$  when  $t_1=1$  and  $t_2= \{1, 2, \dots, 10\}$ . Left:  $\rho=0$ , right:  $\rho=0.5$ .

#### 5.1.2 Steady curve

The methods are compared for  $\rho=\{0, 0.5\}$ , using the steady delay curve.

Both when X and Y are independent and when they have a positive covariance, T2 has longer CED than both the LR methods (except for  $\rho=0.5, t_1=10, t_2=1$ ). Also, both for  $\rho=0$  and  $\rho=0.5$ , LRpar and LRjoint are more similar than LRjoint and T2 (or LRpar and T2). However, for  $\rho=0.5$ , LRpar has a shorter delay for simultaneous changes ( $\tau_X=\tau_Y$ ), compared to LRjoint.

For  $\rho=0$ , the  $CED(t,t)$  is much smaller than  $CED(t,\infty)$  for all three methods. But for  $\rho=0.5$ , when we use LRjoint or T2, the  $CED(t,\infty)$  is only just smaller than  $CED(t,t)$ , since here  $CED(t,t)$  is comparatively long.

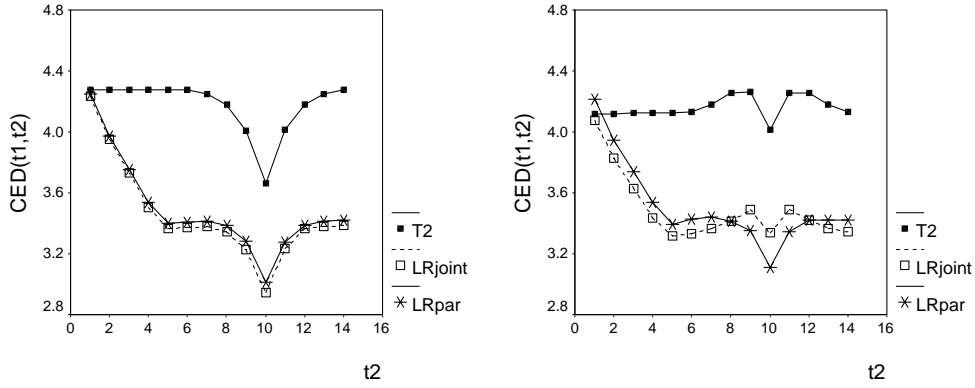


Figure 10:  $CED(t_1, t_2)$  when  $t_1=10$  and  $t_2 = \{1, 2, \dots, 14\}$ . Left:  $\rho=0$ , right:  $\rho=0.5$ .

### 5.2 Effect of change times

The effect of the covariance between  $X$  and  $Y$  is studied,  $\rho = \{0, 0.5\}$ , for each of the three methods. For LRjoint and T2, the  $CED(t,t)$  is shorter than  $CED(t, |t-1|)$ , but the difference is small for  $\rho=0.5$ , compared to  $\rho=0$ . Both T2 and LRjoint do include the components of the Mahalanobis distance, namely  $(x-\mu^D)$ ,  $(y-\mu^D)$  and  $(-2\rho(x-\mu^D)(y-\mu^D))$ . Wessman (1999) pointed out that, when  $\rho > 0.5$ , the Mahalanobis distance is smaller for unit shifts in both processes than for a shift in only one process. For LRjoint and T2, the  $CED(t,\infty)$  is shorter when  $\rho > 0$ , indicating that far apart changes are easy to detect for methods that incorporate  $\rho$ .

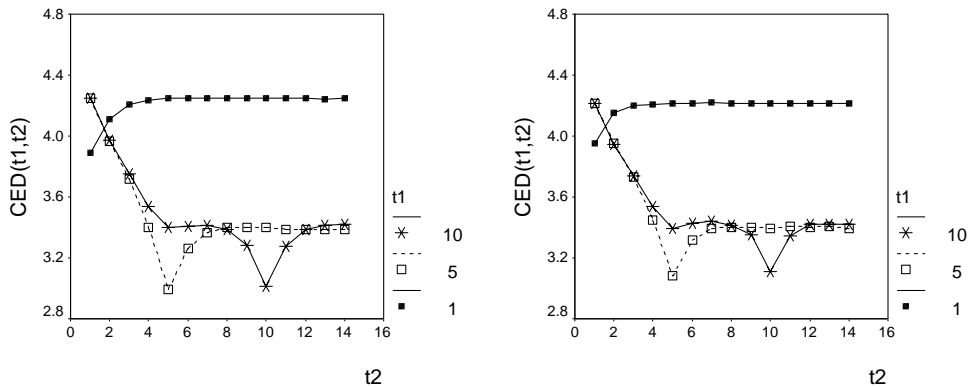


Figure 11a:  $CED(t_1, t_2)$  for LRpar when  $t_1 = \{1, 5, 10\}$ . Left:  $\rho=0.0$ , right:  $\rho=0.5$ .

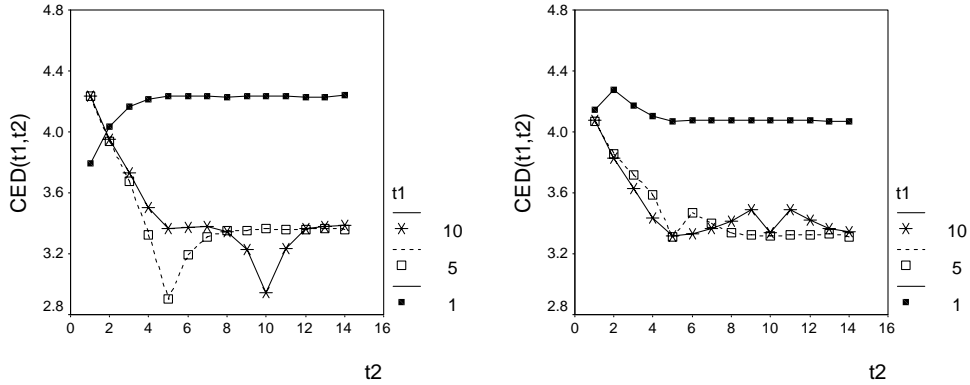


Figure 11b:  $CED(t_1, t_2)$  for LRjoint when  $t_1 = \{1, 5, 10\}$ . Left:  $\rho = 0.0$ , right:  $\rho = 0.0$ .

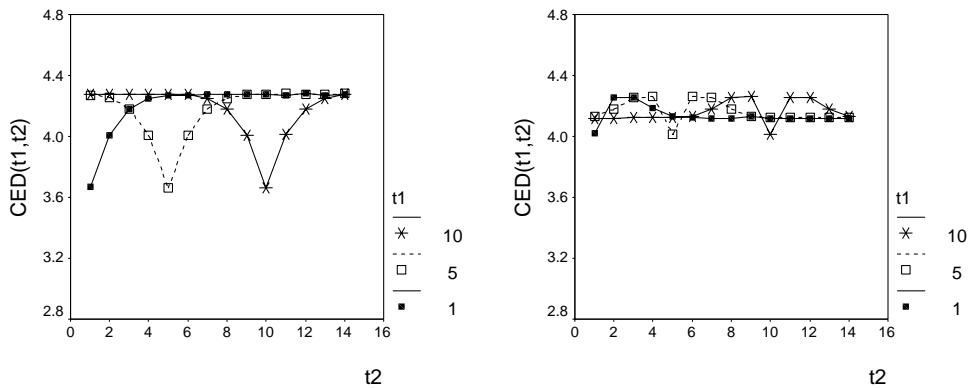


Figure 11c:  $CED(t_1, t_2)$  for T2 when  $t_1 = \{1, 5, 10\}$ . Left:  $\rho = 0.0$ , right:  $\rho = 0.0$ .

A positive covariance between the processes (X and Y) does not affect the CED of LRpar to any large extent.

For LRjoint, the immediate change situation,  $\tau_X = \tau_Y = 1$ , is quicker detected if the processes are independent. For positive  $\rho$ -values X and Y tend to be alike, thus independent processes contain more information. Also for T2, the ( $\tau_X = \tau_Y = 1$ ) situation is easier to detect if the processes are independent.

For late first changes ( $\tau_{(1)} \geq 5$ ), the difference in  $CED(t, \infty)$  between LRpar and LRjoint is larger when  $\rho = 0.5$ , compared to  $\rho = 0$ .

The steady curve of T2 only depends on the distance  $(t_1 - t_2)$ . This was proven generally in Andersson (2005).

### 5.2.1 The delay for close changes

When  $\rho > 0$  (and  $\psi = 0$ ) and the changes occur simultaneously ( $\tau_X = \tau_Y$ ), LRpar has the shortest CED, contrary to when  $\rho = 0$ . These results (Figure 9, 10) indicate that the alarm probability, when the changes do occur close, is low for methods that incorporate  $\rho$  (LRjoint and T2). For LRjoint and T2, the shape of the alarm region changes when  $\rho$  changes, whereas for LRpar the shape is the same (see Section 2.4). A consequence is

that the situation that produces an alarm differs between  $\rho=0$  and  $\rho=0.5$ , for LRjoint and T2.

For LRjoint, one reason for the long delay for ( $\tau_X = \tau_Y$ ) is the assumption of  $\psi=0$ , which results in a low weight for the C-C component. However, the result with long delay for close changes is present also for T2, which does not use information about  $\psi$ .

For both LRjoint and T2, Figure 2 demonstrates that, when  $\rho=0$ , an alarm is called when either  $X'=(X-\mu^D)$  or  $Y'=(Y-\mu^D)$  or  $(X'+Y')$  is large. The sum of two variables that changed simultaneously (from 0 to a positive value  $\mu$ ) is stochastically larger than the sum of two variables with different change times (and thus different expected values, at least at one time point). This can be illustrated by the probability calculations below.

$$P(X'+Y' > k \mid \mu^X=\mu^Y=\mu) = P(Z > (k - (2\mu)) / \sqrt{2\sigma^2}), \text{ for } \tau_X = \tau_Y,$$

$$P(X'+Y' > k \mid \mu^X=0, \mu^Y=\mu) = P\{Z > (k - (\mu)) / \sqrt{2\sigma^2}\}, \text{ for } \tau_X \neq \tau_Y,$$

where the probability is larger for  $\tau_X = \tau_Y$ , i.e.  $P(Z > k-2\mu) > P(Z > k-\mu)$ .

When  $\rho > 0$ , Figure 2 indicate that an alarm is called when the absolute difference,  $|X'-Y'|$ , is large enough. The difference between two variables that change simultaneously (from 0 to a positive value  $\mu$ ) is stochastically smaller than the difference between variables with different change times, illustrated by

$$P(X'-Y' > k \mid \mu^X=\mu^Y=\mu) = P\{Z > (k - (2\mu)) / \sqrt{2\sigma^2(1-\rho)}\}, \text{ for } \tau_X = \tau_Y,$$

$$P(X'-Y' > k \mid \mu^X=\mu, \mu^Y=0) = P\{Z > (k - (\mu)) / \sqrt{2\sigma^2(1-\rho)}\}, \text{ for } \tau_X \neq \tau_Y,$$

where the probability is larger for  $\tau_X \neq \tau_Y$ , i.e.  $P(Z > k-\mu) > P(Z > k-2\mu)$ .

However, for “moderate” covariance,  $\rho=0.5$ , the long CED for simultaneous changes does not always hold. E.g in Figure 11b-11c, we have  $CED(1,1) < CED(1,2)$ . The CED also depends on the model for  $\mu^D$  and  $\mu^C$ . The Mahalanobis distance is proportional to

$$(x(t) - \mu^D)^2 + (y(t) - \mu^D)^2 - 2\rho(x(t) - \mu^D)(y(t) - \mu^D),$$

where  $E[X(t)]$  and  $E[Y(t)]$  depend on  $\tau_X$  and  $\tau_Y$ . We compare the expected Mahalanobis distance, here denoted EM, for two situations: ( $\tau_X = \tau_Y = 1$ ) and ( $\tau_X = 1, \tau_Y = 2$ ).

$$EM(1,1) = (\mu^{C1} - \mu^D)^2 \cdot (1 - 2\rho), \text{ when } \tau_X = \tau_Y = 1,$$

$$EM(1,2) = (\mu^{C1} - \mu^D)^2 + (\mu^{C2} - \mu^D)^2 - 2\rho(\mu^{C1} - \mu^D)(\mu^{C2} - \mu^D), \text{ when } \tau_X = 1, \tau_Y = 2.$$

The inequality  $EM(1,1) < EM(1,2)$  is an indication of the inequality  $CED(1,1) > CED(1,2)$ . The difference  $\{EM(1,1) - EM(1,2)\}$  depends on  $\rho$  as well as  $\{(\mu^{C1} - \mu^D) - (\mu^{C2} - \mu^D)\}$ . If  $\{(\mu^{C1} - \mu^D) - (\mu^{C2} - \mu^D)\}$  is small, then  $\rho$  needs to be close to 1.00 in order for  $EM(1,1) < EM(1,2)$ .

## 6. Discussion

Warning systems are used in many areas: public health, bio terrorism, radiation, pregnancy, intensive care patients. Very often, several processes can be used to detect an underlying change, e.g. confirmed and suspected cases of a disease. Then we need a warning (or surveillance) system for multivariate data. When monitoring more than one process, we must consider the dependency between the processes, the correlation between the change times and the possible differences between the processes.

In this paper, the effect of different types of dependencies is investigated for three methods of multivariate surveillance. One method uses the Hotellings T2, where the multivariate data is reduced to a scalar at each time point. The second method is a parallel system, where the likelihood ratio method is applied to each marginal process, here called the LRpar system. The third method is optimal according to the Shiryaev criterion. It is a simultaneous solution, derived from the joint likelihood ratio, here called LRjoint.

The alarm limits of each of the three systems are adjusted to yield the same false alarm probability for some different situations. One situation is where we assume that the changes occur independent (i.e. change times  $\tau_X$  and  $\tau_Y$  are uncorrelated) and also that the processes (X and Y) are independent, conditional on  $\tau_X$  and  $\tau_Y$ . In another situation, we assume a positive correlation between  $\tau_X$  and  $\tau_Y$  ( $\psi > 0$ , changes tend to occur more simultaneously). In a third situation there is a positive covariance between the monitored processes X and Y, conditional on  $\tau_X$  and  $\tau_Y$  ( $\rho > 0$ ). The evaluation is made using the delay of an alarm, in relation to the first change time.

The methods are first compared for independent processes and no correlation between the change times ( $\rho = 0$ ,  $\psi = 0$ ). For immediate changes in both processes ( $\tau_X = \tau_Y = 1$ ), T2 has the shortest delay, followed by LRjoint and then LRpar. T2 allocates the alarms early, whereas the two LR methods have very few early alarms. LRjoint has a slightly shorter delay than LRpar, since LRjoint is based on three components, corresponding to a change in either X or Y or both, whereas LRpar is based on only two components: change in either X or Y and thus not optimal for simultaneous changes. However, the difference is very small since  $\psi = 0$  (we do not expect simultaneous changes to be too frequent). Also for later simultaneous changes, LRjoint is slightly better than LRpar, whereas T2 here yields a long delay (an effect of allocating the alarms early). Thus, even if T2 is a reduction, which is sufficient at simultaneous change times, the method is not necessarily optimal here. In univariate surveillance, the Shewhart method has a constant CED(t) over all values of t, and the corresponding holds for T2 (constant CED(t,t) over all t). However, for  $\tau > 1$ , Shewhart does not have the shortest delay, and neither has T2. Both Shewhart (univariate) and T2 (multivariate) allocate many alarms early and few at later time points, thus producing longer delay for later changes. For ( $\rho = 0$ ,  $\psi = 0$ ), when we consider different change times, LRpar has the same delay as LRjoint and T2 has the longest delay, because of the alarm allocation.

Next we investigate the effect of assuming a positive correlation between the change times, but still with independent processes ( $\psi = 0.8$ ,  $\rho = 0$ ). For immediate change in both processes,  $\tau_X = \tau_Y = 1$ , LRjoint has the shortest delay, followed by T2 and then LRpar. LRjoint uses the information that the change times are likely to occur simultaneously, which results in a short delay. T2 benefits from allocating the alarm early, whereas LRpar has neither of these advantages. For later simultaneous changes, LRjoint has the shortest

delay. T2, as for  $\psi=0$ , only works well for early simultaneous changes and for late simultaneous changes, LRpar has shorter delay than T2. For  $(\rho=0, \psi=0.8)$ , when we have different change times, LRpar gives slightly shorter delay than LRjoint, and T2 gives the longest delay. LRjoint uses the information that  $\psi>0$ , which is a disadvantage when the changes actually do occur at different times.

Next, the three methods are compared for the situation when X and Y have a positive covariance, but the change times are uncorrelated ( $\rho=0.5, \psi=0$ ). For immediate changes in both processes,  $\tau_X=\tau_Y=1$ , LRpar has the shortest delay, followed by T2 and then LRjoint. For late simultaneous changes, LRpar gives shorter delay than LRjoint (although LRjoint now works better than T2). One explanation to the long delay for LRjoint is the assumption that  $\tau_X$  and  $\tau_Y$  are independent ( $\psi=0$ ). If we used the assumption  $\psi>0$ , LRjoint would perform better at the simultaneous changes. However, the long delay for close changes is present also for T2, which does not use information about  $\psi$ . The alarm statistics of LRjoint and T2 both include  $\rho$ . When  $\rho>0$ , an alarm is called when the absolute difference between  $(X-\mu^D)$  and  $(Y-\mu^D)$  is large, both for LRjoint and for T2. The probability that this difference is large, is lower when  $\tau_X=\tau_Y$ . Both the LRjoint statistic and the T2 statistic includes the Mahalanobis function, thus when the components  $(x-\mu^D)$  and  $(y-\mu^D)$  both are large and  $\rho$  is large, then  $(-2\rho(x-\mu^D)(y-\mu^D))$  is also large and therefore the alarm statistic produces a small value if  $\rho$  is large. Heuristically, we may say that when X and Y have a positive covariance, they give similar information, especially if they have the same distribution ( $\tau_X=\tau_Y$ ) and this may explain the long delay. For  $(\rho=0.5, \psi=0)$ , when we have different change times, LRjoint gives a shorter delay than LRpar and T2 yields the longest delay.

Generally, the difference in delay is large between LRjoint and T2, in so that LRjoint gives shorter delay than T2, except at early, simultaneous changes. The difference between LRjoint and LRpar is smaller than between LRjoint and T2.

If all processes change immediately, T2 gives shortest delay, in case of total independence ( $\rho=0, \psi=0$ ), whereas LRjoint is best if there is a positive correlation between the change times ( $\rho=0, \psi>0$ ). If there is a positive covariance between the processes and no correlation between the change times ( $\rho>0, \psi=0$ ), then LRpar gives the shortest delay for  $\tau_X=\tau_Y=1$ . For later simultaneous changes, LRjoint is best, for all non-negative correlations, when the processes are independent ( $\psi\geq 0, \rho=0$ ). For processes with a positive covariance, LRpar yields the shortest delay for later simultaneous changes, at least when no correlation is assumed between the change times ( $\rho>0, \psi=0$ ).

If the processes change at different time points, LRjoint gives the shortest delay, except for a positive correlation between the change times ( $\psi>0$ ). This holds for both independent processes and when the covariance is positive ( $\rho\geq 0$ ).

In this paper we only deal with positive dependency. A negative correlation between the change times,  $\psi<0$ , implies that the change times do not coincide. Then it would be best to have a very small weight for the component for simultaneous changes. If the processes themselves have a negative covariance ( $\rho<0$ ), the implication is that  $X=-Y$ . Then the alarm region would constitute of large values of  $(X+Y)$  and the alarm region for  $\rho<0$  would be similar to that of a positive correlation between the change times ( $\psi>0$ ).



## Appendix

The  $M_{ij}(t)$  is based on two distances, namely  
 $d_i(t)=(\mu C_i(t)-\mu D)/\sigma$  and  $d_j(t)=(\mu C_j(t)-\mu D)/\sigma$ .

, we compare  $M_{1,1}(t)$  and  $M_{1,2}(t)$  (below the time index  $t$  is dropped):

$$M_{1,1} = \frac{d_1^2 + d_1^2 - 2\rho d_1 d_1}{(1-\rho^2)} \text{ and } M_{1,2} = \frac{d_1^2 + d_2^2 - 2\rho d_1 d_2}{(1-\rho^2)} .$$

If the inequality  $M_{1,2} > M_{1,1}$  holds, this does indicate that  $CED(1,1) < CED(1,2)$ . The inequality between the  $M$ -values holds when  $(d_1^2 - d_2^2) / 2d_1(d_1 - d_2) < \rho$ . Thus, if  $d_2$  is small compared to  $d_1$ , then  $M_{1,1} < M_{1,2}$  for rather moderate covariance. But if  $d_2$  is closer in size to  $d_1$ , then  $\rho$  needs to be close to 1.00 in order for  $M_{1,1} < M_{1,2}$ . In our model, the relative difference between  $d_1$  and  $d_2$  is in fact rather small, thus only for large values of  $\rho$ , we would get  $CED(1,1) > CED(1,2)$ .

## References

- Abu-Shawiesh, M. and Abdullah, M. (2001) A new robust bivariate control chart for location. *Communications in Statistics-Simulation and Computation*, **30**, 513-530.
- Alt, F. B. (1985) *Multivariate quality control*, Wiley.
- Andersson, E. (2004) The impact of intensity in surveillance of cyclical processes. *Communications in Statistics-Simulation and Computation*, **33**, 889-913.
- Andersson, E. (2005) The T2 method in multivariate on-line surveillance. The delay of an alarm. *Working paper*, Statistical Research Unit, Göteborg University, Sweden.
- Andersson, E., Bock, D. and Frisé, M. (2004) Detection of turning points in business cycles. *Journal of Business Cycle Measurement and Analysis*, **1**, 93-108.
- Aparisi, F., Jabaloyes, J. and Carrión, A. (2001) Generalized variance chart design with adaptive sample sizes. The bivariate case. *Communications in Statistics-Simulation and Computation*, **30**, 931-949.
- Baron, M. (2002) Bayes and asymptotically pointwise optimal stopping rules for the detection of influenza epidemics. In *Case Studies in Bayesian Statistics*, Vol. 6 (Eds, Gatsonis, C., Kass, R. E., Carriquiry, A., Gelman, A., Higdon, D., Pauler, D. K. and Verdini, I.) Springer-Verlag, New York, pp. 153-163.
- Bock, D. (2003) Similarities and differences between statistical surveillance and certain decision rules in finance. *Research report 2003:4*, Statistical Research Unit, Göteborg University, Sweden.
- Bodden, K. M. and Rigdon, S. E. (1999) A program for approximating the in-control ARL for the MEWMA chart. *Journal of Quality Technology*, **31**, 120-123.
- Charnes, J. M. (1995) Tests for special causes with multivariate autocorrelated data. *Computers & Operations Research*, **22**, 443-453.
- Cheng, A. Y. and Liu, R. Y. (2000) Monitoring multivariate aviation safety data by data depth: Control charts and threshold systems. *IIE Transactions*, **32**, 861-872.
- Crosier, R. B. (1988) Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics*, **30**, 291-303.
- Does, R. J. M. M., Roes, K. C. B. and Trip, A. (1999) Handling multivariate problems with univariate control charts. *Journal of Chemometrics*, **13**, 353-369.
- Frisé, M. (2003) Statistical surveillance. Optimality and methods. *International Statistical Review*, **71**, 403-434.
- Frisé, M. and de Maré, J. (1991) Optimal Surveillance. *Biometrika*, **78**, 271-80.
- Frisé, M. and Sonesson, C. (2002) Optimal surveillance based on exponentially weighted moving average. *Research report 2002:1*, Department of Statistics, Göteborg University, Sweden.
- Frisé, M. and Wessman, P. (1999) Evaluations of likelihood ratio methods for surveillance. Differences and robustness. *Communications in Statistics. Simulations and Computations*, **28**, 597-622.
- Gan, F. (1993) An optimal-design of EWMA control charts based on median run-length. *Journal of Statistical Computation and Simulation*, **45**, 169-184.
- Gan, F. F. (1997) Joint monitoring of process mean and variance. *Nonlinear Analysis*, **30**, 4017-4024.

- Guerrero-Cusumano, J.-L. (1995) Testing variability in multivariate quality control: A conditional entropy measure approach. *Information Sciences*, **86**, 179-202.
- Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**, 357-384.
- Hawkins, D. M. (1991) Multivariate Quality Control Based on Regression-Adjusted Variables. *Technometrics*, **33**, 61-75.
- Hotelling, H. (1947) Multivariate Quality Control. In *Techniques of statistical analysis*(Eds, Eisenhart , C., Hastay, M. W. and Wallis, W. A.) McGraw-Hill, New York.
- Javaheri, A. and Houshmand, A. A. (2001) Average run length comparison of multivariate control charts. *Journal of Statistical Computation and Simulation*, **69**, 125-140.
- Jolayemi, J. K. (2000) A model for the statistical design of multivariate control charts with multiple control regions. *Applied Mathematics and Computation*, **109**, 73-91.
- Järpe, E. (1999) Surveillance of spatial patterns. *Communications in Statistics. Theory and Methods*, **28**, 3009-3025.
- Kalagonda, A. A. and Kulkarni, S. R. (2003) Diagnosis of Multivariate Control Chart Signal Based on Dummy Variable Regression Technique. *Communications in Statistics-Theory and Methods*, **32**, 1665-1684.
- Kang, L. and Albin, S. L. (2000) On-line monitoring when the process yields a linear profile. *Journal of Quality Technology*, **32**, 418-426.
- Koskinen, L. and Öller, L.-E. (2004) A Classifying Procedure for Signalling Turning Points. *Journal of Forecasting*, **23**, 197-214.
- Kourti, T. and MacGregor, J. F. (1996) Multivariate SPC methods for process and product monitoring. *Journal of Quality Technology*, **28**, 409-428.
- Love, T. E. and Linderman, K. (2003) A Weibull process failure mechanism for the economic design of MEWMA control charts. *Journal of Statistical Computation and Simulation*, **73**, 195-202.
- Lowry, C. A., Woodall, W. H., Champ, C. W. and Rigdon, S. E. (1992) A multivariate exponentially weighted moving average control chart. *Technometrics*, **34**, 46-53.
- Lu, X. S., Xie, M., Goh, T. N. and Lai, C. D. (1998) Control chart for multivariate attribute processes. *International Journal of Production Research*, **36**, 3477-3489.
- Marshall, A. W. and Olkin, I. (1997) A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, **84**, 641-652.
- Mason, R. L., Chou, Y. M., Sullivan, J. H., Stoumbos, Z. G. and Young, J. C. (2003) Systematic patterns in T-2 charts. *Journal of Quality Technology*, **35**, 47-58.
- Mason, R. L., Tracy, N. D. and Young, J. C. (1995) Decomposition of T2 for multivariate control chart interpretation. *Journal of Quality Technology*, **27**, 99-108.
- Molnau, W. E., Runger, G. C., Montgomery, D. C., Skinner, K. R., Lored, E. N. and Prabhu, S. S. (2001) A program for ARL calculation for multivariate EWMA charts. *Journal of Quality Technology*, **33**, 515-521.
- Neftci, S. (1982) Optimal prediction of cyclical downturns. *Journal of Economic Dynamics and Control*, **4**, 225-41.
- Page, E. S. (1954) Continuous inspection schemes. *Biometrika*, **41**, 100-114.

- Petzold, M., Sonesson, C., Bergman, E. and Kieler, H. (2004) Surveillance in longitudinal models. Detection of intra-uterine growth restriction. *Biometrics*, **60**, 1025-1033.
- Pignatiello, J. J. and Runger, G. C. (1990) Comparisons of Multivariate CUSUM Charts. *Journal of Quality Technology*, **22**, 173-186.
- Qiu, P. H. and Hawkins, D. (2001) A rank-based multivariate CUSUM procedure. *Technometrics*, **43**, 120-132.
- Roberts, S. W. (1959) Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, **1**, 239-250.
- Royston, P. (1991) Identifying the fertile phase of the human menstrual cycle. *Statistics in Medicine*, **10**, 221-240.
- Runger, G. C., Keats, J. B., Montgomery, D. C. and Scranton, R. D. (1999) Improving the performance of the multivariate exponentially weighted moving average control chart. *Quality and Reliability Engineering Internato*, **15**, 161-166.
- Sepúlveda, A. and Nachlas, J. A. (1997) A simulation approach to multivariate quality control. *Computers & Industrial Engineering*, **33**, 113-116.
- Serel, D. A., Moskowitz, H. and Tang, J. (2000) Univariate (X)over-bar control charts for individual characteristics in a multinormal model. *Iie Transactions*, **32**, 1115-1125.
- Shewhart, W. A. (1931) *Economic Control of Quality of Manufactured Product*, MacMillan and Co., London.
- Shiryayev, A. N. (1963) On optimum methods in quickest detection problems. *Theory of Probability and its Applications.*, **8**, 22-46.
- Sonesson, C. and Frisén, M. (2005) Multivariate surveillance. In *Spatial surveillance for public health*(Eds, Lawson, A. and Kleinman, K.) Wiley.
- Stoumbos, Z. G. and Jones, L. A. (2000) On the properties and design of individuals control charts based on simplicial depth. *Nonlinear Studies*, **7**, 147-178.
- Stoumbos, Z. G. and Sullivan, J. H. (2002) Robustness to non-normality of the multivariate EWMA control chart. *Journal of Quality Technology*, **34**, 260-276.
- Sun, K. and Basu, A. P. (1995) A characterization of a bivariate geometric distribution. *Statistics & Probability Letters*, **23**, 307-311.
- Talluri, S. and Sarkis, J. (2002) A methodology for monitoring system performance. *International Journal of Production Research*, **40**, 1567-1582.
- Wessman, P. (1998) Some Principles for surveillance adopted for multivariate processes with a common change point. *Communications in Statistics. Theory and Methods*, **27**, 1143-1161.
- Wessman, P. (1999) The surveillance of several processes with different change points. *Research report*, Department of Statistics, Göteborg University, Sweden.
- Wikström, C., Albano, C., Eriksson, L., Fridén, H., Johansson, E., Nordahl, Å., Rännar, S., Sandberg, M., Kettaneh-Wold, N. and Wold, S. (1998) Multivariate process and quality monitoring applied to an electrolysis process: Part I. Process supervision with multivariate control charts. *Chemometrics and Intelligent Laboratory Systems*, **42**, 221-231.
- Woodall, W. H. and Ncube, M. M. (1985) Multivariate Cusum Quality Control Procedures. *Technometrics*, **27**, 285-292.

Yumin, L. (1996) An improvement for mewma in multivariate process control\*1.  
*Computers & Industrial Engineering*, **31**, 779-781.

## Research Report

- |         |   |  |
|---------|---|--|
| 2003:1  | Holgersson, T. & Shukur, G.:            | Testing for multivariate heteroscedasticity.   |
| 2003:2  | Holgersson, T.:                         | Testing for multivariate autocorrelation.  |
| 2003:3  | Petzold, M. & Sonesson, C.:             | Detection of intrauterine growth restriction.  |
| 2003:4  | Bock, D.:                               | Similarities and differences between statistical surveillance and certain decision rules in finance.                           |
| 2003:5  | Holgersson, T. & Lindström, F.:         | A comparison of conditioned versus unconditioned forecasts of the VAR(1) process.  |
| 2003:6  | Bock, D.:                               | Early warnings for turns in business cycles and finance.   |
| 2003:7  | Lindström, F.:                          | On prediction accuracy of the first order vector auto regressive process.  |
| 2003:8  | Petzold, M. & Jonsson, R.:              | Maximum Likelihood Ratio based small-sample tests for random coefficients in linear regression.                                |
| 2003:9  | Petzold, M.:                            | Preliminary testing in a class of simple non-linear mixed models to improve estimation accuracy.                               |
| 2003:10 | Frisén, M. & Gottlow, M.:               | Graphical evaluation of statistical surveillance.  |
| 2003:11 | Frisén, M.:                             | Statistical measures for evaluation of methods for syndromic surveillance.   |
| 2003:12 | Andersson, E.:                          | A monitoring system for detecting starts and declines of influenza epidemics.  |
| 2004:1  | Bock, D., van Dijk, D. & Franses, P.H.: | Monitoring macroeconomic volatility.   |
| 2004:2  | Bock, D.                                | Aspects on the control of false alarms in statistical surveillance and the impact on the return of financial decision systems. |